

Assignment 3 Part 2

Part 2 has landed! We will continue with the Model as service part!

Things you need to do: (Week 2)

- 1) **Build a Docker Container**
- 2) **Schedule Airflow**

Note:

- 1) This is an extension of Assignment 3 so use the same repo
- 2) Make sure there are no errors in your python files when you submit it.
- 3) Have a requirement.txt file in your repository so we can install packages which you have used.

Quick Flashback:

- 1) Last week we gave you a new dataset and you had built a Machine learning model on it.
- 2) You had an API which makes inferences to the model
- 3) You deployed the API (Using Heroku or some other cool platform) and made a frontend for your users!

That's it!

This week we will repeat step 3 from above Flashback but will deploy it with docker and build an inference pipeline with Airflow!

Building Dockers

In this module you will build a docker container for your Fast API codebase.

Docker useful links:

- 1) <https://docker-curriculum.com>
- 2) <https://stackify.com/docker-tutorial/>
- 3) <https://www.youtube.com/watch?v=77nUeNZ2954>

Here you build the docker of your Fast API code base, push the image to any docker registry and deploy the container on any cloud platform. (Suggest Google cloud).

Suggested workflows: using GitHub actions to build a docker file! (Team4,5 have been already doing this)

Submission:

- 1) Codebase for Docker file.
- 2) Running FAST Api application

Airflow

While training the model we did not use testing data. In this module we will build an airflow DAG which will take the testing data as input and do the inference of all images sequentially.

Aim: You must make a DAG which will take all testing images and output the confusion matrix of the data. Schedule this DAG on airflow with any time you like and make a report of the workflows.

Like how many times did the DAG run in a day, how many times it failed/success, what was the time taken to run the DAG etc.

You need to show the confusion matrix from all runs on streamlit and the analysis. Make a new tab airflow analysis. Keep it very user friendly and readable.

Note: We know here that every time confusion matrix will be same as we will be passing the same image in everyrun.

Useful Links:

- 1) Confusion Matrix: <https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>
- 2) <https://airflow.apache.org/docs/apache-airflow/stable/tutorial.html>
- 3) <https://towardsdatascience.com/a-complete-introduction-to-apache-airflow-b7e238a33df>
- 4) <https://medium.com/abn-amro-developer/data-pipeline-orchestration-on-steroids-apache-airflow-tutorial-part-1-87361905db6d>

Submission:

- 1) *The airflow folder with DAGS in it.*
- 2) *Screenshot of your airflow dashboard*

Streamlit:

Keep your streamlit from last assignment and add the airflow output (Confusion Matrix, report) in it!

This is due this week 23rd July 2022 11:59 AM

Any doubts? Message on slack or teams